

# On The Effect of Mutations in Bovine or Camel Chymosin on the Thermodynamics of Binding $\kappa$ -Caseins

Samiul M. Ansari<sup>1</sup>, Jesper Sørensen<sup>2</sup>, Birgit Schiøtt<sup>2</sup>, and David S. Palmer<sup>\*1</sup>

<sup>1</sup>*Department of Pure and Applied Chemistry, University of Strathclyde, Thomas Graham Building, 295  
Cathedral Street, Glasgow, G1 1XL, Scotland*

<sup>2</sup>*Interdisciplinary Nanoscience Center (iNANO) and Department of Chemistry, University of Aarhus,  
Langelandsgade 140, DK-8000 Aarhus, Denmark*

<sup>\*</sup>*Corresponding Author. Email: david.palmer@strath.ac.uk. Tel: +44 (0) 141 548 4178*

Keywords: alanine scanning, binding thermodynamics, MM-3DRISM, cheese manufacturing,  
dairy industry, aspartic protease, enzyme, molecular dynamics

Short title: Chymosin– $\kappa$ -Casein Complexes

## Abstract

Bovine and camel chymosins are aspartic proteases that are used in dairy food manufacturing. Both enzymes catalyse proteolysis of a milk protein,  $\kappa$ -casein, which helps to initiate milk coagulation. Surprisingly, camel chymosin shows a 70 % higher clotting activity than bovine chymosin for bovine milk, while exhibiting only 20 % of the unspecific proteolytic activity. By contrast, bovine chymosin is a poor coagulant for camel milk. Although both enzymes are marketed commercially, the disparity in their catalytic activity is not yet well understood at a molecular level, due in part to a lack of atomistic resolution data about the chymosin -  $\kappa$ -casein complexes. Here, we report computational alanine scanning calculations of all four chymosin -  $\kappa$ -casein complexes, allowing us to elucidate the influence that individual residues have on binding thermodynamics. Of the 12 sequence differences in the binding sites of bovine and camel chymosin, eight are shown to be particularly important for understanding differences in the binding thermodynamics (Asp112Glu, Lys221Val, Gln242Arg, Gln278Lys, Glu290Asp, His292Asn, Gln294Glu, and Lys295Leu. Residue in bovine chymosin written first). The relative binding free energies of single-point mutants of chymosin are calculated using the molecular mechanics three dimensional reference interaction site model (MM-3DRISM). Visualisation of the solvent density functions calculated by 3DRISM reveals the difference in solvation of the binding sites of chymosin mutants.

## 1 Introduction

Bovine chymosin is an aspartic protease enzyme which selectively cleaves  $\kappa$ -casein proteins in milk, initiating coagulation to aid digestion.[1] It has recently been found that camel chymosin is an effective clotting agent for bovine milk.[2] The camel variant shows a 70% higher clotting activity for bovine milk, while exhibiting only 20% of the unspecific proteolytic activity.[2] By contrast, bovine chymosin poorly coagulates camel milk. Both bovine and camel chymosin are now marketed for use in the food industry, which has led to renewed interest both in understanding the disparity in their catalytic behaviour and in designing new variants with improved milk-clotting

properties.

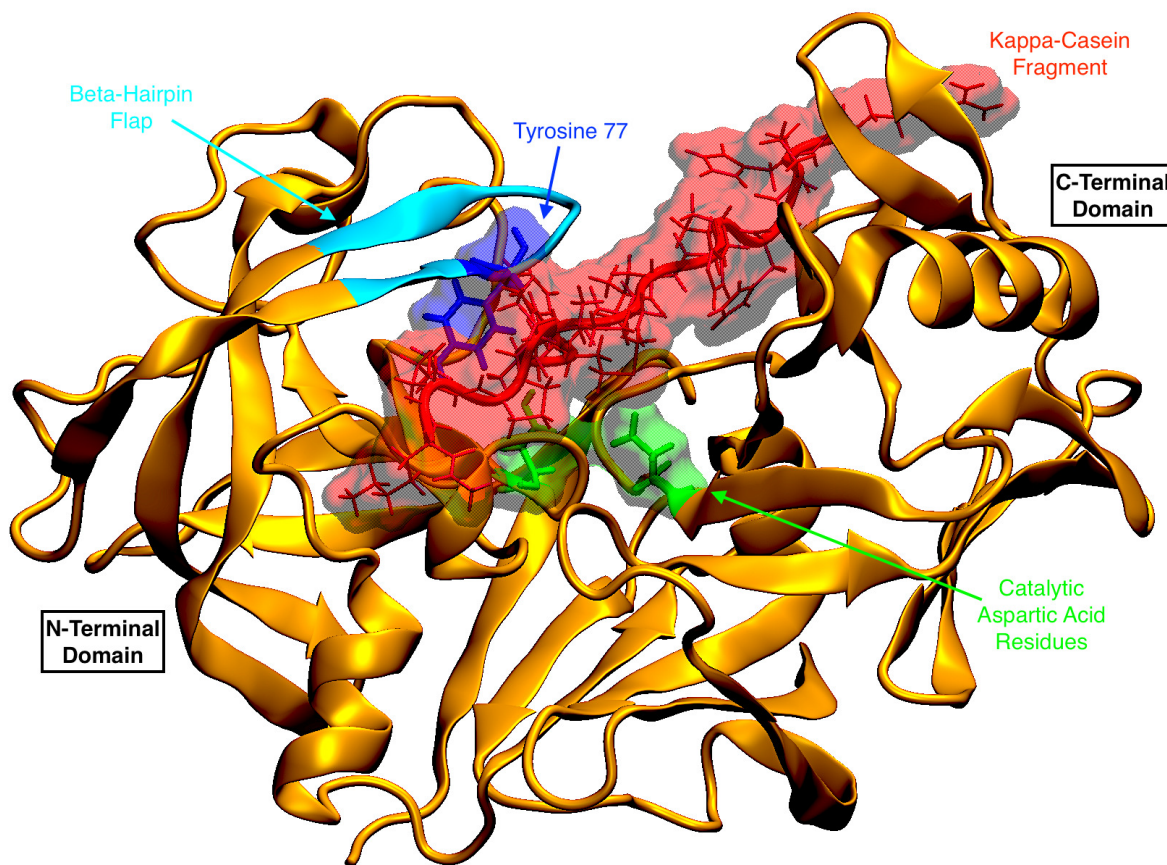


Figure 1: Depiction of bovine chymosin–bovine  $\kappa$ -casein complex.  $\kappa$ -casein fragment in red aligned across the binding cleft of chymosin. Catalytic aspartic acid residues in green located within the binding cleft.

Bovine and camel chymosin have high sequence similarity (94%) and identity (85%) and similar three-dimensional structures. They both comprise 323 amino acids that fold into a pseudosymmetric bi-lobal structure forming a central binding cleft containing the catalytic residues Asp34 and Asp216. The side chains of the catalytic aspartic acid residues extend towards each other in a planar geometry,[3] which is stabilised by a network of hydrogen bonds with two threonine residues, referred to as “*the fireman’s grip*”.[4] Similar features are found in other homologous aspartic proteases.[5, 6, 7, 8]

Within the substrate-binding cleft, there are 12 differences in the primary structure of bovine

and camel chymosin. Both variants of the enzyme have a positively charged patch on the N-terminal (residues 50-60) and a negatively charged patch on the C-terminal (residues 240-260) that interact with  $\kappa$ -casein.[9]. The positive patch is larger in camel chymosin through the replacement of a Gln56 in bovine chymosin by His56 in camel. The negatively charged patch in camel chymosin is found to be less negative through the replacement of Asp249 and Asp251 in bovine chymosin by Asn249 and Gly251 in the camel variant. In camel chymosin there are two additional positively charged patches that are not found in bovine chymosin.[10] The first resides in the C-terminal where a small positive patch comprises residues Arg242, Arg254 and Lys278, and the corresponding residues in bovine chymosin are hydrophilic but neutral. The second is found at the base of the binding cleft where the residues are Arg150 and Arg316 in camel chymosin, but Gln150 and Leu316 in bovine chymosin.

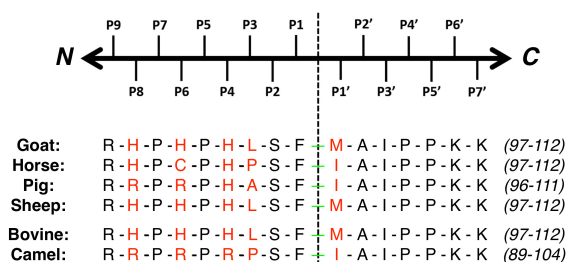


Figure 2: The aligned primary sequence of the chymosin sensitive region of  $\kappa$ -casein of different species. The Pn and Pn' numbering follows the Schechter and Berger nomenclature,[11] where n increases with the distance from the scissile bond. Residues that differ between some of the species are highlighted in red. The residue numbers are shown to the right in parenthesis.

The substrate of chymosin is the 169 residue milk protein,  $\kappa$ -casein, which is found on the surface of milk serum aggregates called *casein micelles*. [12] The hydrophilic region of  $\kappa$ -casein protrudes from the micelles giving the structure stability against spontaneous aggregation. [13] Chymosin selectively cleaves bovine  $\kappa$ -casein at the PheP1-MetP1' bond and camel  $\kappa$ -casein at the PheP1-IleP1' bond, causing the hydrophilic C-terminal end of  $\kappa$ -casein to dissociate, thereby destabilizing the casein micelles initiating the release of insoluble casein proteins which results in milk clotting. [13] (The Px or Px' nomenclature is used to describe  $\kappa$ -casein residues on the two sides of the cleavage site, e.g. Ser104, Phe105, Met106 and Ala107 in bovine  $\kappa$ -casein are referred

to as SerP2, PheP1, MetP1' and AlaP2', respectively. Regions of chymosin that interact with the P2, P1, P1' and P2' residues are denoted S2, S1, S1' and S2' pockets, respectively.[11]). The amino acid sequences of  $\kappa$ -caseins from different species in the region of the cleavage site are given in Figure 2.

The crystal structure of a chymosin-inhibitor complex (RCSB PDB ID: 1CZI) and previous molecular modelling studies, suggest that  $\kappa$ -casein binds to chymosin in an extended secondary structure.[13, 14] This is supported by circular dichroism, solution NMR and molecular modelling studies of unbound  $\kappa$ -casein, all showing an extended structure in the region of the scissile bond.[15, 16] Molecular modelling and mutagenesis studies, propose that the P8-P7' residues are located in the chymosin binding cleft during catalysis. Recently the arginine residue in the P9 position has been implicated in binding because it is conserved in bovine, camel, pig, buffalo and goat chymosin.[17] Furthermore, a ArgP9His mutant is observed to be a poor substrate. [18] In  $\kappa$ -casein, residue SerP2 appears to be essential for the catalysis to take place.[19] The hydrophobic residues LeuP3, AlaP2' and IleP3' are crucial in giving the structure its hydrophobic qualities.[20] In camel  $\kappa$ -casein LeuP3 is replaced with a hydrophobic proline residue, retaining the same hydrophobic qualities as bovine chymosin. At the typical pH of milk, 6.6 - 6.7, micelles carry a net negative charge.

In 1995 a study of complexes involving the HisP8-LysP6' fragment from bovine  $\kappa$ -casein coupled with bovine chymosin and porcine pepsin was investigated through restrained molecular dynamics.[14] The authors initially suggested that a cis-peptide bond between HisP8-His99 was crucial for interactions between Asp247 and His98. However in 1997 they revised this theory on the basis of results from molecular dynamics simulations using a longer peptide chain.[15] The 1997 study reported favourable interactions between  $\kappa$ -casein and chymosin (HisP4:Glu245, HisP6:Asp297 and LysP6':Glu133 respectively).

In 2010, unrestrained molecular dynamics simulations were carried out on bovine chymosin complexed with a fragment of  $\kappa$ -casein (ArgP9-LysP7').[3] The trajectories showed the substrate binds in an extended pose and charged residues flank the scissile bond, stabilising the binding pose. The  $\kappa$ -casein fragment can be seen binding to both N- and C-terminal domains: residues LysP6'

and LysP7' displace a conserved water molecule to bind to the N-terminal domain and the HPHPH sequence in residues P8-P4 binds to the C-terminal domain.

A solvent binding and computational alanine scanning study of chymosin– $\kappa$ -casein complexes in 2013 highlighted that water binding sites on the surface of bovine chymosin help to stabilize the complex.[21] The authors demonstrated that relative binding thermodynamics of single-point mutants in bovine chymosin–bovine  $\kappa$ -casein complexes can be accurately calculated using molecular integral equation theory techniques.

Here we investigate the importance of individual amino acid residues in chymosin by calculating the influence each residue has on the binding free energy of chymosin to its substrate,  $\kappa$ -casein. We focus on the 12 residues that are different in the active sites of bovine and camel chymosin in order to elucidate the influence they have on binding in each complex. For each of these residues, computational alanine scanning calculations are performed in all four chymosin– $\kappa$ -casein complexes (Bov/Bov, Bov/Cam, Cam/Bov, Cam/Cam) using the Molecular Mechanics Three-Dimensional Reference Interaction Site Model (MM-3DRISM) methodology. Using MM-3DRISM permits accurate estimates of solvation and binding free energies and allows us to investigate solvent density effects that could not be studied by implicit continuum solvent models (as in e.g. MM-PBSA).

## 2 Theory

### 2.1 Calculation of $\Delta G_{bind}$

The thermodynamic parameter that characterizes the binding of a ligand (L) by a receptor (R) is the binding free energy ( $\Delta G_{bind}$ ) for the process:[22]



Methods such as thermodynamic integration (TI) and free energy perturbation (FEP) are considered thermodynamically rigorous and in principle are accurate for calculating relative binding free

energies between two equilibrium states. However both methods require a great deal of simulation time to provide adequate sampling, making them unsuitable for large scale studies. To tackle this practical issue, many different end-point techniques have been developed to predict binding free energies at lower computational expense i.e. the linear-interaction-energy (LIE) approach,[23] and the closely related Molecular Mechanics Generalised Born Surface Area (MMGBSA), Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA),[24, 25] and MM-3DRISM methods. Where FEP and TI spend the majority of simulation time investigating intermediate states, end-point techniques investigate just two (bound and unbound) states, significantly reducing computational cost. The solvent environment strongly influences the binding free energy, modulating competitive solvent binding effects and hydrophobicity.[26] Here we employ the MM-3DRISM method, which uses a statistical mechanics based solvent model to provide a realistic model of molecular solvation effects. MM-3DRISM affords accurate estimates of binding free energy and has previously been used in modelling a wide-variety of protein-ligand complexes. [27, 21]

In MM-3DRISM, the binding free energy is computed according to:

$$\Delta G_{bind} = G_{solvated}(complex) - G_{solvated}(enzyme) - G_{solvated}(substrate) \quad (2)$$

The free energies of each species is evaluated as:

$$G_{solvated} = \langle E_{gas} \rangle + \langle \Delta G_{hyd} \rangle - TS \quad (3)$$

$$\langle E_{gas} \rangle = \langle E_{internal} \rangle + \langle E_{electrostatic} \rangle + \langle E_{vdW} \rangle \quad (4)$$

$$\langle E_{internal} \rangle = \langle E_{bond} \rangle + \langle E_{angle} \rangle + \langle E_{torsion} \rangle \quad (5)$$

$E_{gas}$  describes the average energy of a species in the gas phase as a sum of internal, electrostatic ( $E_{electrostatic}$ ) and van der Waals ( $E_{vdW}$ ) energy contributions, obtained through a molecular

mechanics forcefield.  $E_{bond}$ ,  $E_{angle}$  and  $E_{torsion}$  contribute to the internal energy  $E_{internal}$  through the strain caused by deviation of bonds, angles, and torsions from their equilibrium values.  $\Delta G_{hyd}$  describes the hydration free energy and is computed by the 3D-RISM calculation. Entropic contributions were calculated from rotational, translational and vibrational contributions, with the latter computed by normal mode analysis.[28] The binding free energy of a single complex is calculated through the average binding free energy from a set of different conformations of the protein-ligand complex. We calculate  $\Delta G_{hyd}$  using the advanced pressure correction (PC+) free energy functional. The PC+ functional contains no empirical parameters and has been shown to give accurate predictions of hydration free energies for neutral and ionised solutes, in both pure water and salt solutions at a wide-range of temperatures.[29, 30, 31, 32, 33] As a comparison, we also present results for three other free energy functionals: Partial-Series-Expansion-3 (PSE-3), Pressure Correction (PC) and Gaussian Fluctuations (GF). For the calculation of relative solvation free energies, as is required in computational alanine scanning, we show that similar results are obtained using the PSE-3, PC and PC+ functionals because the differences between these functionals partially cancel out. All thermodynamic data that could not be reported in the main article due to space limitations is provided in the Supporting Information.

## 2.2 Computational Alanine Scanning

The main aim of this work is to elucidate the effect that differences in the primary sequence of bovine and camel chymosin have on binding thermodynamics. It is therefore useful to have an estimate of the contribution that a single residue makes to the total binding free energy. For protein complexes, computational alanine scanning is one of the most widely used techniques. The difference between the mutant and wild-type complex is calculated, where one or more residues in the wild-type complex has been mutated to an alanine residue, described in Equation 6.

$$\Delta\Delta G_{bind} = \Delta G_{bind,mutant} - \Delta G_{bind,wildtype} \quad (6)$$



Here, we probe the effects on binding thermodynamics by single point mutations on natural mutants between bovine and camel chymosin in the various chymosin- $\kappa$ -casein complexes through computational alanine scanning.

### 2.3 Calculation of $\Delta G_{hyd}$ – 3D-RISM Solvent Model

3DRISM is a statistical mechanics based method for modelling biomacromolecules in solution phases. [34, 35, 36, 37] Only a brief description is given here as detailed explanations have been published elsewhere.[38] The 3DRISM equation relates 3D intermolecular *solvent site - solute* total correlation functions ( $h_\alpha(\mathbf{r})$ ) to direct correlation functions ( $c_\alpha(\mathbf{r})$ ) (where  $\alpha$  corresponds to the solvent sites): [36, 34]

$$h_\alpha(\mathbf{r}) = \sum_{\xi=1}^{N_{solvent}} \int_{R^3} c_\xi(\mathbf{r} - \mathbf{r}') \chi_{\xi\alpha}(|\mathbf{r}'|) d\mathbf{r}' \quad (7)$$

Here  $N_{solvent}$  is the number of sites in a solvent molecule. The term  $\chi_{\xi\alpha}(r)$  represents the bulk solvent susceptibility function which describes the mutual correlations of sites  $\xi$  and  $\alpha$  of solvent molecules in the bulk solvent. It is obtained from the solvent intramolecular correlation function ( $\omega_{\xi\alpha}^{solv}(r)$ ), solvent site number density ( $\rho_\alpha$ ) and the site-site radial total correlation functions ( $h_{\xi\alpha}^{solv}(r)$ ):  $\chi_{\xi\alpha}(r) = \omega_{\xi\alpha}^{solv}(r) + \rho_\alpha h_{\xi\alpha}^{solv}(r)$  (we imply that each site in the molecule is unique, therefore  $\rho_\alpha = \rho$  for all  $\alpha$ ).[36] In this study these functions were obtained by solving the 1D-RISM equations of the solvent.[36, 39]

$N_{solvent}$  closure relations are introduced in order to calculate  $h_\alpha(\mathbf{r})$  and  $c_\alpha(\mathbf{r})$ :

$$h_\alpha(\mathbf{r}) = \exp(-\beta u_\alpha(\mathbf{r}) + h_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r}) + B_\alpha(\mathbf{r})) - 1 \quad (8)$$

$$\alpha = 1, \dots, N_{solvent}$$

Here  $B_\alpha(\mathbf{r})$  are bridge functionals,  $u_\alpha(\mathbf{r})$  represents the 3D interaction potential between the solute molecule and  $\alpha$  solvent site, and  $\beta = 1/k_B T$ , where  $k_B$  is the Boltzmann constant and  $T$  is the temperature. The 3D interaction potential between the  $\alpha$  site of solvent and solute molecule, is

predicted as a superposition of the site-site interaction potentials, depending on only the absolute distance between the particular solvent site and solute site, described in Equation 8. The long-range electrostatic interaction term and the short-range term (Lennard-Jones potential) is used to represent the site-site interaction potential.

The exact bridge functionals  $B_\alpha(\mathbf{r})$  used in Equation 8 are an infinite series of integrals over high order correlation functions, and therefore incomputable. In practice, approximations are incorporated[40, 41, 36] or the form of these functionals are estimated from molecular simulations.[42] The simplest approximate bridge functional is the hyper-netted chain (HNC) closure in which  $B_\alpha(\mathbf{r}) = 0$ . [32] Less severe approximations that also provide better numerical convergence of the 3D-RISM equations are offered by the PSE-1, PSE-2 and PSE-3 closures,[43] the first of which is more commonly referred to as the Kovalenko and Hirata (KH) closure. [44] Here the final solutions to the 3DRISM equations are obtained using the PSE-3 closure as described in more detail in the Methods section.

Within 3D-RISM theory there are a number of approximate functionals that allows one to use the total  $h_\alpha(\mathbf{r})$  and direct  $c_\alpha(\mathbf{r})$  correlation functions to analytically obtain values of the hydration free energy (HFE).[45, 27, 46]

### 2.3.1 Partial Series Expansion-3 free energy functional (3D-RISM(PSE-3))

For the PSE-3 functional, the solute's excess chemical potential (solvation free energy) at infinite dilution is derived from the 3D-RISM solute-solvent correlation functions as follows (Equation 9):

$$\Delta G_{hyd}^{PSE-3} = \Delta G_{hyd}^{HNC} - k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_\alpha \int_V \left[ \Theta[h_\alpha(\mathbf{r})] \frac{\Xi_\alpha(\mathbf{r})^{n+1}}{(n+1)!} \right] d\mathbf{r} \quad (9)$$

where  $\rho_\alpha$  is the number density of solvent sites  $\alpha$ ,  $\Theta$  is a Heaviside step function, and  $\Delta G_{hyd}^{HNC}$  is the solvation free energy calculated using the hypernetted-chain functional, which is given by: [31]

$$\Delta G_{hyd}^{HNC} = k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_\alpha \int_V \left[ \frac{1}{2} h_\alpha^2(\mathbf{r}) - \frac{1}{2} h_\alpha(\mathbf{r}) c_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r}) \right] d\mathbf{r} \quad (10)$$

### 2.3.2 Gaussian Fluctuations free energy functional (3D-RISM(GF))

Developed by Chandler, Singh and Richardson, for 1D-RISM, and adopted by Kovalenko and Hirata for the 3D-RISM case[36, 47], the Gaussian fluctuations (GF) HFE functional is given as:

$$\Delta G_{hyd}^{GF} = k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_{\alpha} \int_{R^3} \left[ -\frac{1}{2} c_{\alpha}(\mathbf{r}) h_{\alpha}(\mathbf{r}) - c_{\alpha}(\mathbf{r}) \right] d\mathbf{r} \quad (11)$$

### 2.3.3 Pressure Corrected free energy functional

**3D-RISM(PC)** The PC free energy functional is designed as an improvement on the standard 3DRISM solvation free energy functionals that over estimate the solvent pressure. To counteract this the PC functional subtracts all mechanical work required to create the cavity ( $P\Delta V$ ) from the hydration free energy ( $\Delta G_{Hyd}^{3D-RISM}$ ), as shown in Equation 12.[48, 32]

$$\Delta G_{Hyd}^{PC} = \Delta G_{Hyd}^{3D-RISM} - P\Delta V \quad (12)$$

Here  $P$  represents the 3D-RISM pressure and  $\Delta V$  represents the volume change of the system upon solvation.  $P\Delta V$  is computed using methods described by Misin *et al.*[31, 32]  $\Delta G_{Hyd}^{PC}$  simply refers to the pressure corrected hydration free energy, 3D-RISM(PC).

**3D-RISM(PC+)** The PC+ free energy functional is a further improvement on the PC functional where just the non-ideal mechanical work is subtracted from the hydration free energy. To accomplish this the ideal gas pressure expansion,  $P_{id}$  is used to represent the ideal mechanical work,  $P_{id}\Delta V$  and is added to Equation 12.

$$\Delta G_{Hyd}^{PC+} = \Delta G_{Hyd}^{3D-RISM} - P\Delta V + P_{id}\Delta V \quad (13)$$

Here, and in the formula for the PC functional,  $\Delta G_{Hyd}^{3D-RISM}$  is the solvation free energy calculated using the PSE-3 free energy functional (Equation 9). Although there is no compelling explanation as to why PC+ performs better than PC, there have been numerous reports of its benefits in the

literature.[49, 32, 50, 31, 48] The PC+ functional has been shown to give accurate predictions of hydration free energies for neutral and ionised solutes, in both pure water and salt solutions at a wide-range of temperatures.[29, 30, 31, 32, 33] It has also been successfully applied to the prediction of solvation free energies in organic solvents. [49]

## 3 Materials and Methods

### 3.1 Molecular Dynamics Simulations

A total of 100 ns of unrestrained molecular dynamics simulations were run for each of the four chymosin-P9-P7' -  $\kappa$ -casein complexes (Bov/Bov, Bov/Cam, Cam/Bov,Cam/Cam). Input coordinates for each complex were taken from our previous work.[3, 51]

The molecular dynamic simulations were run with the AMBER FF03 force field parameters developed by Duan *et al.* using NAMD[52] and the TIP3P water model.[53] Production simulations were run in the isothermal-isobaric (NPT) ensemble[54] at 1 atm. Langevin dynamics maintained the system temperature at 300K and the pressure was regulated by the Nosé-Hoover Langevin piston pressure control [55], the piston was set to a target of 1.01325 bar, period set to 200fs, and decay set to 100 fs.[56] Periodic boundary conditions were applied to each system and the electrostatic interactions were calculated using the particle mesh Ewald (PME) method.[57, 58, 59] Van der Waals interactions had a cut-off distance of 10 Å and a switching distance of 9 Å. All hydrogen to hetero-atom bond distances were constrained by the SHAKE algorithm.[60, 61] The velocity verlet algorithm was set to update the equations of motions every 2 fs, and snapshots were stored every 2 ps.

Each simulation system was equilibrated for 4 ns. A 96 ns simulation was generated with a snapshot every 400 ps providing a 240 frame trajectory for analysis. To reduce unnecessary computational expense, MM-3DRISM and normal-mode entropy calculations were carried out on every third frame of this trajectory, as per previous MM-3DRISM studies. [27, 21]

### 3.2 MM-3DRISM Calculations

Binding free energy of the  $\kappa$ -casein fragment to chymosin was calculated using the MM-3DRISM[27, 62] method at 298.15 K as implemented in AmberTools15, [63, 39] using a locally modified version of the MMPBSA.py program in which we implemented the PC and PC+ free energy functionals.[49] The calculations were carried out on single trajectories of each complex, this has proven to be both computationally more efficient and provides results closer to experimental values through cancellation of errors.[64] All interactions were computed with the AMBER FF03 force field. 3D-RISM calculations were performed with the assumption of an infinitely dilute solute. Solvent was modelled using a modified SPC water model (as implemented in the AmberTools package) with a water density of 55.343 mol/l. The modified SPC water model was used to avoid numerical convergence issues.[65] The buffer parameter was set to give a minimum distance of 18 Å between the solute and the edge of the solvent box. The calculations employed the MDIIS iterative scheme,[66], where 5 MDIIS vectors were used, and a MDIIS step size of 0.7. The grid spacing was 0.75 Å. Solvent susceptibility functions required as input to the 3D-RISM calculations were calculated with the dielectrically consistent 1D-RISM. The grid spacing for 1D functions was 0.025 Å, which gave a total of 16 384 grid points. We employed the MDIIS iterative scheme, where we used 20 MDIIS vectors, an MDIIS step size of 0.3, and a residual tolerance of  $10^{-12}$ . The solvent was considered to be pure water with a number density  $0.0333 \text{ Å}^3$  and a dielectric constant of 78.497.

### 3.3 Computational Alanine Scanning

Computational alanine scanning calculations were carried out for 12 residues in the binding site that are natural mutants (different amino acids) in bovine and camel chymosin. All of these residues were within 4 Å of a residue in  $\kappa$ -casein for at least 70% of each of the molecular dynamics trajectories (measured using custom written VMD scripts)(Figure 3).

The Massova and Kollman[67] protocol was used to carry out the alanine scanning calculations, employing the same trajectories used in the binding free energy calculations. The Massova and Kollman method assumes that the mutations do not significantly change the dynamics of the enzyme-

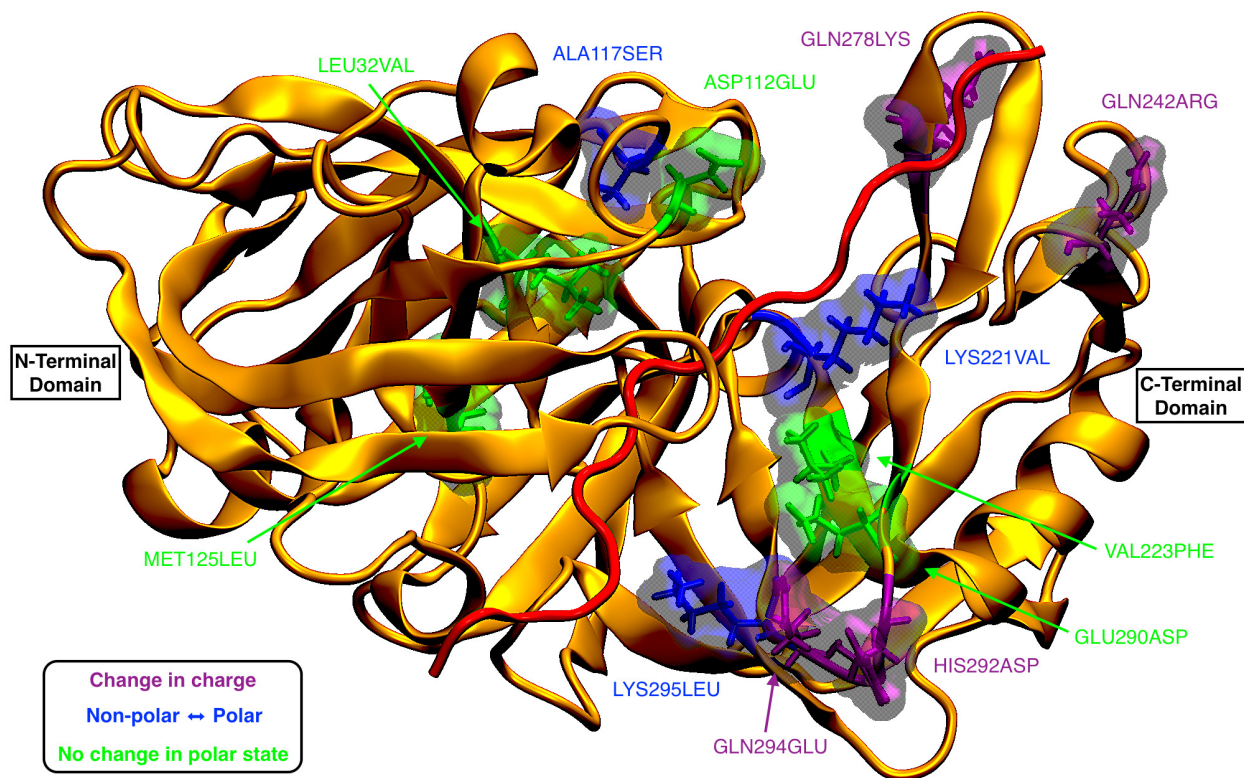


Figure 3: Depiction of residues that are naturally different between bovine and camel chymosin (*natural mutants*), shown on a bovine chymosin–bovine  $\kappa$ -casein complex. The labels refer to the bovine residue first and the camel counterpart second. Purple residues show that the amino acid has a different polar charge in each version of chymosin. Blue residues show the amino acid has gone from a polar to a non-polar one, or *vice-versa*. Green residues show that the polarity (and charge) remains the same in camel and bovine chymosin even though the amino acids are different.

substrate system, thereby introducing the mutation after the simulations have been performed should provide accurate results. This assumption has been shown to be valid for a wide-range of protein-ligand systems.[68] It is also supported by molecular dynamics simulations of bovine chymosin complexes (unbound, inhibitor bound and substrate bound) which shows binding incurs no significant change in the conformation of the protein backbone.[69, 13, 3] The protocol has a number of advantages including being computationally much less demanding, and most importantly the use of the same trajectories allows for the cancellation of errors, resulting in more accurate results.[67, 70] In accordance with previous studies[67, 70, 71] and limited by computational expense, the entropy term was neglected for all alanine scanning calculations, since using the

Massova and Kollman protocol the difference in entropy between mutant and wildtype proteins is expected to be negligible.

To calculate the difference that an alanine mutation makes on binding free energy we subtract the wild-type binding free energy from the mutant to give  $\Delta\Delta G_{bind}$ , ( $\Delta\Delta G_{bind} = \Delta G_{mutant} - \Delta G_{wildtype}$ ). A negative  $\Delta\Delta G_{bind}$  indicates a favourable mutation (the native system has higher binding free energy than the mutant), and positive results are unfavourable mutations (the native system has a lower binding free energy compared to the mutant). As reported in previous work[21, 72] residues can be classified on a basis of magnitude of  $\Delta\Delta G_{bind}$ ; warm ( $\geq 1$  kcal/mol) or hot-spots ( $\geq 2$  kcal/mol), representing a disproportionate contribution to the binding free energy.

## 4 Results and Discussion

### 4.1 Binding Free Energies

The binding free energies of the wildtype chymosin- $\kappa$ -casein complexes were calculated using four different solvation free energy functionals in MM-3DRISM. Extensive previous benchmarking on solvation free energy data of organic molecules indicates that the PC+ functional gives more accurate results than the GF, PSE-3 or PC functionals.[29, 30, 31, 32, 33] In this section, which discusses binding free energies, we therefore focus on the results obtained using the PC+ functional, while the results of the other functionals are provided in the Supporting Information. The calculated binding free energies must be interpreted with caution because they do not include some terms relating to the loss of conformational freedom on binding (due to the use of a single-trajectory approach to the MM-3DRISM calculations) and because they only include harmonic contributions to the vibrational entropy; neither of these problems unduly affect the computational alanine scanning results because of favourable cancellation of errors. Table 1 presents enthalpic and entropic components of the free energy calculated by MM-3DRISM(PC+). The results indicate that binding is thermodynamically favourable for all four complexes, but that the native complexes exhibit more favourable binding than the cross-complexes. For all four complexes, a large favorable change in the gas phase contribution

to the binding free energy ( $\Delta G_{gas}$ ) is opposed by an unfavorable change in the hydration free energy ( $\Delta G_{hyd}$ ) of similar magnitude. The entropic contributions to the binding free energy are of similar magnitude for all complexes. The binding free energies calculated by MM-3DRISM(PC+) for the Bov/Bov and Cam/Bov systems are observed to be in good general agreement with those obtained from MM-PBSA by Sorensen *et al.* (Table 1). The binding free energies obtained by these two methods differ by  $\approx 2$  kcal/mol in both cases, which is not negligible, but is surprisingly consistent given the size of the peptide substrates and the difficulties associated with predicting absolute binding free energies from molecular simulation. Unfortunately, it is impossible to compare either of these sets of results to experiment, since neither the binding free energies nor the hydration free energies of the protein-ligand complexes considered here have been reported.

Table 1: Different Components of Binding Free Energy Calculated for the Various Chymosin- $\kappa$ -Casein Complexed using MM-3DRISM(PC+) Methodology. MM-PBSA results taken from reports by Sørensen *et al.*[72] All Values Given are in units of kcal/mol.

<b>Energy</b>	<b>CAM/CAM</b>		<b>CAM/BOV</b>		<b>BOV/CAM</b>		<b>BOV/BOV</b>	
	<b>Mean</b>	<b>SE<sup>a</sup></b>	<b>Mean</b>	<b>SE</b>	<b>Mean</b>	<b>SE</b>	<b>Mean</b>	<b>SE</b>
$\Delta G_{gas}^b$	-1330.1	9.8	-1029.6	6.6	-1595.4	7.2	-1388.0	9.9
$\Delta G_{hyd}^c$	1211.7	9.1	932.4	6.1	1494.7	6.8	1277.3	9.1
$\Delta G_{total}^d$	-118.4	1.4	-97.2	1.2	-100.7	1.1	-110.8	1.4
$T\Delta S^e$	-76.5	1.5	-61.9	1.3	-68.1	1.4	-70.2	1.1
$\Delta G_{bind}^f$	-41.9	2.1	-35.3	1.8	-32.6	1.8	-40.6	1.8
$\Delta G_{bind}^{MM-PBSA}$ (Ref. [72]) <sup>g</sup>	-	-	-33.4	0.8	-	-	-42.8	0.7

<sup>a</sup> Standard Error; <sup>b</sup> Total Gas Phase Free Energy; <sup>c</sup> Total Hydration Free Energy; <sup>d</sup> Total Energy; <sup>e</sup> Total Entropy; <sup>f</sup> Total Binding Free Energy. <sup>g</sup> Results taken from reports by Sørensen *et al.*[72]

## 4.2 Alanine Scanning

To determine the importance of individual residues for free energy of binding, alanine scanning calculations have been performed in all four complexes (Bov/Bov, Bov/Cam, Cam/Bov and Cam/Cam), from 96 ns MD simulations of each complex. The alanine scanning results that reveal a significant difference for a given amino acid position in the four complexes will be grouped into two classes, corresponding to whether it is the native residue in bovine or in camel chymosin that contributes



more favourably to the binding free energy. A negative  $\Delta\Delta G$  shows that mutating the natural residue to alanine will result in a stronger binding, and a positive  $\Delta\Delta G$  means the alanine mutation will result in a weaker binding.

#### 4.2.1 Favoured Native Camel Residues

**LYS221VAL** Residue 221 lies in the S4 pocket in close contact with either HisP4 (bovine) or ArgP4 (camel) of  $\kappa$ -casein.[72] On the basis of binding free energy calculations of the wildtype bovine complex only, it has previously been suggested that Val221 should be more favoured than Lys221 for the binding of bovine  $\kappa$ -casein.[72] The results presented in Figure 4 are in good agreement with that prediction. Figure 4 shows that the Lys221Ala mutation in bovine chymosin is favoured (because it reduces unfavourable polar interactions with HisP4 in bovine  $\kappa$ -casein), whereas by contrast the Val221Ala mutation in camel chymosin is unfavoured (because it reduces favourable nonpolar interactions). For the camel  $\kappa$ -casein substrate (5), the same trend is observed, but the effects are greater because ArgP4 (camel) is larger and more basic than HisP4 (bovine). The unfavourable interaction between ArgP4 and Lys221 correlates with the observation that camel  $\kappa$ -casein is a poor substrate for bovine chymosin. This is supported by the experimental observation that a LysP4 mutant of bovine  $\kappa$ -casein is also a poor substrate for bovine chymosin.[73]

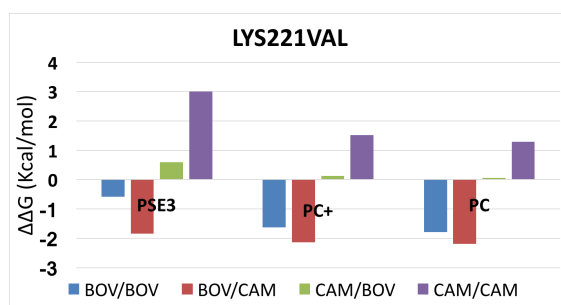


Figure 4: Comparison of alanine scanning results of residue 221 (Lys221 in bovine chymosin, Val221 in camel chymosin), on the four different chymosin– $\kappa$ -casein complexes with three different MM-3DRISM calculation methods. A negative  $\Delta\Delta G$  represents a favourable mutation, and positive results are unfavourable.

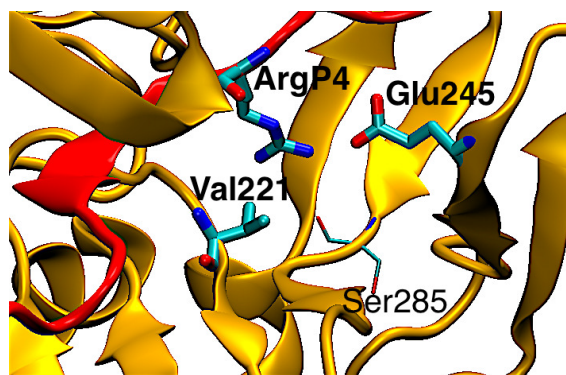


Figure 5: Snapshot of Val221 and ArgP4 (in bold for clarity) at 75.2 ns of Cam/Cam complex MD simulation. Residue ArgP4 interacts with both Val221 and Glu245.

**ASP112GLU** Residue 112 (Asp112 in bovine chymosin, Glu112 in camel chymosin) lies at one end of a short  $\alpha$ -helical region of chymosin (residues 112 to 116) near the surface of the binding cleft. The importance of the Asp112Glu mutation to chymosin  $\kappa$ -casein binding thermodynamics is not immediately obvious from crystallographic data since the mutation lies in the N-terminal domain of chymosin, whereas the neighbouring P9-P1 residues of  $\kappa$ -casein bind predominantly to the C-terminal domain. Nonetheless, in all four complexes, mutating residue 112 to alanine is observed to be thermodynamically unfavourable (Figure 6). Furthermore, a clear difference is observed in the values of  $\Delta\Delta G_{bind}$  for alanine scanning in the four complexes. Both the importance of residue 112 and the trend in the alanine scanning results can be partly explained by a salt bridge between residue 112 of chymosin and P8 of  $\kappa$ -casein, which is observed to form for some part of each of the simulations. For example, in the Cam/Cam complex, a relatively stable Glu112-ArgP9 salt bridge is observed throughout the majority of the simulation (Figure 7). Consequently, mutating Glu112 to alanine results in an unfavourable change in binding free energy because of the loss of the salt bridge. By contrast, in the Bov/Cam system, in which Asp112 in chymosin interacts with ArgP8 in  $\kappa$ -casein, the salt bridge is observed less frequently during the molecular dynamics simulation and the loss of binding free energy due to an Asp112Ala mutation is lower. This trend agrees with recent experimental and computational research that suggests that in solution Arg forms weaker salt bridges with Asp than Glu, [74] (although the chymosin environment would be expected to modulate this effect to some extent). However, in the context of chymosin- $\kappa$ -casein

complexes, it may also be partly due to the fact that the sidechain of Glu is longer than Asp and, hence, it can orientate itself better with respect to ArgP8. The alanine scanning data for bovine or camel chymosin binding to bovine  $\kappa$ -casein reveals a similar trend with the stronger salt bridge in the Cam/Bov complex (Glu112-HisP8) giving rise to a slightly larger value of  $\Delta\Delta G_{bind}$  than that in the Bov/Bov complex (Asp112-HisP8. HisP8 was modelled as the charged histidinium ion in agreement with previous work).[3]

The  $\alpha$ -helical region containing Asp112Glu has been implicated in allosteric activation of bovine chymosin by the P8-P4 residues of bovine  $\kappa$ -casein.[8] In this mechanism, the *His-Pro cluster* (HPHPH in P8-P4 of bovine  $\kappa$ -casein) interacts with the  $\alpha$ -helix, which both allows the  $\beta$ -hairpin flap in residues 72-84 of chymosin to twist and causes the side chain of Phe114 to vacate a pocket that is occupied by Tyr77 in the open conformation. The interaction between Asp112 and the P8 residue of  $\kappa$ -casein is therefore a potential target for protein engineering aimed at modifying the self-inhibited to open transition of Tyr77 in the bovine complex. However, further experimental research would be required to verify how these processes occur in the complexes involving camel chymosin or camel  $\kappa$ -casein.

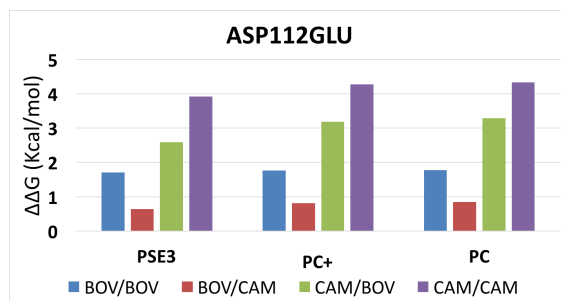


Figure 6: Comparison of alanine scanning results of residue 112 (Asp112 in bovine chymosin, Glu112 in camel chymosin).

**GLU290ASP, HIS292ASN, GLN294GLU and LYS295LEU** Residues 290 to 295 form an unstructured loop region above the centre of the binding cleft in the C-terminal domain, opposite the  $\beta$ -hairpin flap in the N-terminal domain. The loop region is known to be more flexible than the surrounding residues as indicated by the crystallographic B-factors of the backbone atoms, which

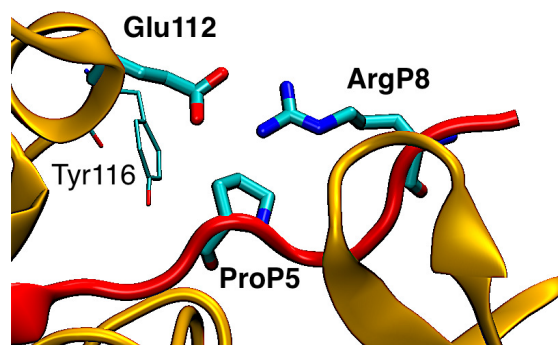


Figure 7: Snapshot of Glu112 and ArgP8 (in bold for clarity) at 18.8 ns of Cam/Cam complex MD simulation.

are  $\sim 40 \text{ \AA}^2$  in the loop compared to an average of  $\sim 21\text{-}22 \text{ \AA}^2$  in the protein.[10] Indeed, in earlier crystallographic studies, residues 291 to 293 were considered to be too flexible to be resolved accurately.[69] The primary sequence of residues 290 to 295 is ENHSQK in bovine chymosin and DNNSEL in camel chymosin, which reveals four mutations: Glu290Asp, His292Asn, Gln294Glu and Lys295Leu. Analysis of the molecular dynamics trajectories shows that the sidechains of residues Glu290Asp, Gln294Glu and Lys295Leu point towards  $\kappa$ -casein in all four complexes, while Ser293 and, to a lesser extent, Asn291 are solvent exposed. His292Asn lies in the most flexible region at the tip of the loop (B-factor  $> 40 \text{ \AA}^2$  in both bovine and camel crystal structures). An ensemble of different conformations are observed throughout the molecular dynamics simulations, but on average Asn292 in camel chymosin is more solvent exposed than His292 in bovine chymosin regardless of the identity of the substrate.

In all four complexes, residue 290 (Glu290 in bovine chymosin, Asp290 in camel chymosin) forms intermittent hydrogen bonds with the sidechain of SerP2 and non-specific interactions with IleP3'; both of these residues are conserved in bovine and camel  $\kappa$ -casein (as well as goat, horse, pig and sheep  $\kappa$ -casein. Figure 2). The alanine scanning calculations show a weak preference for Asp in the 290 position, but mutating either Glu290 or Asp290 to alanine is unfavourable, since it incurs the loss of a hydrogen bond to SerP2 (Figure S6, Supporting Information). A more selective influence on the binding free energy is observed in positions 292, 294 and 295. Mutating His292 to alanine strengthens the binding by  $\sim 2 \text{ kcal/mol}$  in the bovine chymosin complexes, whilst

mutating Asn292 to alanine in the camel chymosin complexes has essentially no effect because the residue is largely solvent exposed (Figure S7, Supporting Information). Similarly, mutating Lys295 to alanine strengthens the binding by  $\sim 2$  kcal/mol in both bovine chymosin complexes (Figure 8), whilst mutating Leu295 to alanine in the camel chymosin complexes has little effect because both Leu295 and Ala295 make similar weak van der Waals interactions with  $\kappa$ -casein (IleP3' and ProP5'). (Figure 9). In the 294 position, there is a weak preference for the Glu294 residue in camel chymosin, but the Gln294 residue in bovine chymosin contributes approximately the same amount to the binding free energy as an Ala residue (Figure S8, Supporting Information). The 294 residue points towards the sidechains of the P1 and P3 residues in  $\kappa$ -casein and is partially solvated in all four complexes.

An additional consideration is that Asn291 is a known glycosylation site in both bovine and camel chymosin.[10] Here we have considered the unglycosylated variants since these predominate in commercial products. However, *A. Niger* fermentation can glycosylate proteins at N $^{\delta 2}$  atoms of Asn in Asn-X-Thr/Ser sequences, of which there are two in both bovine (Asn252 and Asn291) and camel chymosin (Asn100 and Asn291). Glycosylation is favoured at Asn-X-Thr sequences (Asn100 Camel) as compared to Asn-X-Ser sequences (Asn252 and Asn291 bovine, Asn291 camel). Approximately 10% of bovine chymosin produced by *A. Niger* fermentation is glycosylated. A reduction in clotting activity is observed when camel chymosin is glycosylated at Asn291.[10]

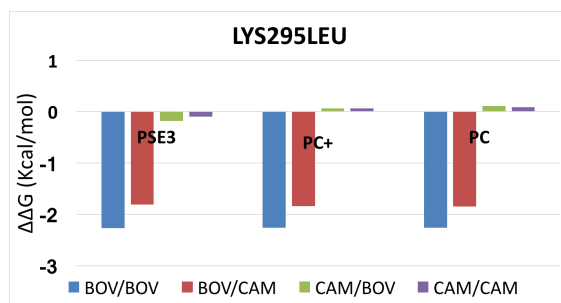


Figure 8: Comparison of alanine scanning results of residue 295 (Lys295 in bovine chymosin, Leu295 in camel chymosin).

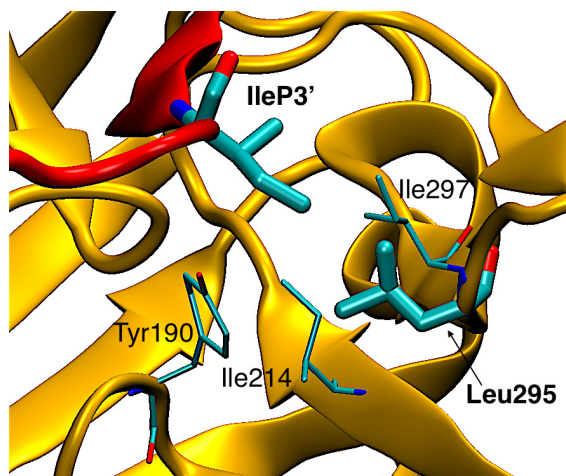


Figure 9: Snapshot of Leu295 and IleP3' at 17.2 ns of Cam/Cam complex MD simulation.

#### 4.2.2 Favoured Native Bovine Residues

**GLN242ARG** Residue 242 (Gln242 in bovine chymosin, Arg242 in camel chymosin) resides in a predominantly uncharged polar region on the surface of the C-terminal domain, where it interacts with the ArgP9 residue of  $\kappa$ -casein. Although early structural studies focused on the P8-P7' residues of  $\kappa$ -casein only,[14] the importance of the P9 position for binding has since been recognised because ArgP9 is conserved in bovine, camel, pig, buffalo, horse, and goat chymosin.[75] Furthermore, a variant of bovine  $\kappa$ -casein, in which the P9 position is occupied by a histidine, has been shown to be a poor substrate for bovine chymosin.[18]. The Gln242Arg mutation is the only sequence difference in the S9 pocket of bovine and camel chymosin.

The alanine scanning results for residue 242 (Figure 10) reveal a significant difference between the bovine and camel variants. For the Bov/Bov and Bov/Cam systems, Gln242 can be seen extending towards ArgP9 throughout the molecular dynamics simulations (Figure 11). Consequently, mutating Gln242 to alanine results in weaker binding in these systems, as indicated by the positive  $\Delta\Delta G$  obtained by alanine scanning (Figure 10). By contrast, in camel chymosin, where the interaction of Arg242 with ArgP9 is electrostatically and sterically unfavourable, the sidechain of Arg242 is observed to extend partly out of the binding pocket (Figure 12). Here mutating Arg242 in camel chymosin to alanine shows a clear improvement in binding free energy with a reduction in  $\Delta\Delta G$  by

~ 3.5 kcal/mol (Figure 10). The results suggest that the Gln residue observed in wildtype bovine enzyme is more favoured for binding than the Arg residue from its camel counterpart. The role of the residues in the S9 pocket has not previously been defined, but it may be to help orientate the neighbouring P8-P4 residues of  $\kappa$ -casein, which have been implicated by both experimental and modelling studies in the allosteric activation of chymosin.[76, 73, 8]

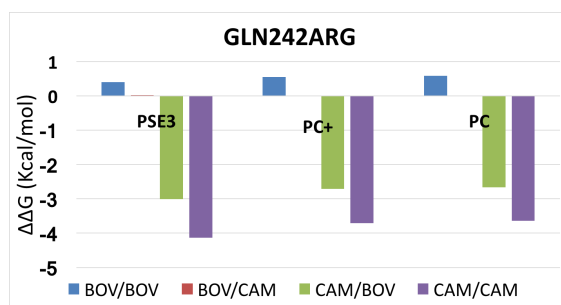


Figure 10: Comparison of alanine scanning results of residue 242 (Gln242 in bovine chymosin, Arg242 in camel chymosin)

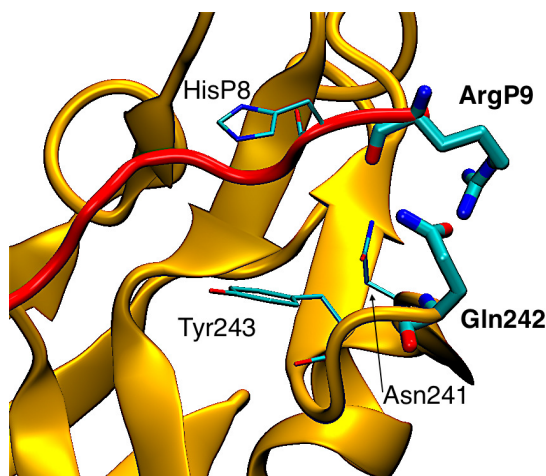


Figure 11: Snapshot of Gln242 and ArgP9 (in bold for clarity) at 15.2 ns of Bov/Bov complex MD simulation. The side chains of both residues extended towards each other.

**GLN278LYS** Residue 278 (Gln278 in bovine chymosin, Lys278 in camel chymosin) resides in a predominantly uncharged region on the surface of the N-terminal domain of chymosin, where it interacts with the P6 residue in  $\kappa$ -casein, which is HisP6 in bovine and the larger and more basic ArgP6 in camel. The S6 pocket is an open cleft formed by the sidechains of the Ser277, Asp279

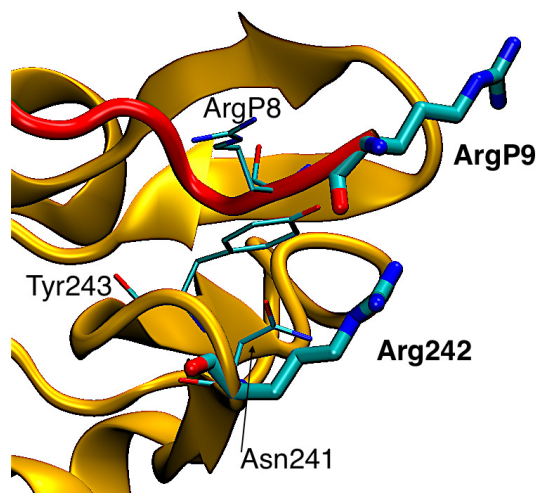


Figure 12: Snapshot of Arg242 and ArgP9 (in bold for clarity) at 53.2 ns of Cam/Cam complex MD simulation. The side chains of both residues extended away from the binding pocket.

and Thr284 residues in the C-terminal domain and Asp13 and Ser14 residues in the N-terminal domain, all of which are conserved in bovine and camel chymosin. The only mutation site near the S6 pocket, Gln278Lys, is not a common target for protein engineering because crystallographic data shows that it lies at the bottom of the open cleft with the Gln or Lys sidechain pointing away from the binding site in apo bovine or camel chymosin, respectively. In solution, as revealed by the molecular dynamics simulation, however, the flexibility of the sidechain of residue 278 allows it to extend over the open cleft of the S6 pocket bringing it closer to the P6 residue of  $\kappa$ -casein. Alanine scanning results for residue 278 show that mutating the natural camel residue to alanine favours binding of bovine or camel  $\kappa$ -casein, as shown by the negative  $\Delta\Delta G$  in Figure 13. Here the mutation to alanine removes an unfavourable Lys-Arg (Cam/Cam) or Lys-His(Cam/Bov) interaction. By contrast, mutating the bovine chymosin residue, Gln278 to alanine shows no significant change in binding free energy for either the Bov/Bov or Bov/Cam complex. The results suggest that Gln or Ala are favoured over Lys in the 278 position of chymosin.



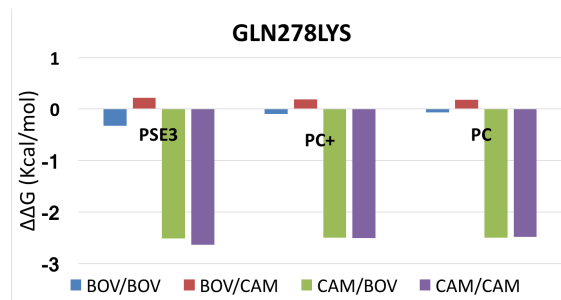


Figure 13: Comparison of alanine scanning results of residue 278 (Gln278 in bovine chymosin, Lys278 in camel chymosin), on the four different chymosin- $\kappa$ -casein complexes with three different MM-3DRISM calculation methods. A negative  $\Delta\Delta G$  represent a favourable mutation, and positive results are unfavourable.

### 4.2.3 Other Residues

The remaining four residues that were analysed through computational alanine scanning show no significant difference in  $\Delta\Delta G_{bind}$  between bovine or camel chymosin complexes (Leu32Val, Ala117Ser, Met125Leu, and Val223Phe). The results for these residues are presented in the Supporting Information. It was found that changes in binding free energy were either too little to be considered a significant change, or the change was the same throughout all systems.

## 5 Changes in Solvent Density Distribution Due to Single-Point Mutations

As well as permitting estimates of solvation thermodynamics, the 3DRISM calculations provide information about the local solvation of protein-ligand binding sites that can be readily visualised. The 3DRISM solvent density functions,  $g_\alpha(\mathbf{r}) = h_\alpha(\mathbf{r}) + 1$ , give the spatial distribution of solvent density on a grid around the protein-ligand complex. The changes in solvent density distribution that occurs from a single-point mutation in a chymosin- $\kappa$ -casein complex can be illustrated by taking the difference between the spacial density distribution functions of the mutant and wildtype complexes for a single configuration of the complex:  $\Delta g(\mathbf{r})_{m/w} = g(\mathbf{r})_{mutant} - g(\mathbf{r})_{wildtype}$ . [21] Figure 14 shows the corresponding isosurfaces at  $\Delta g(\mathbf{r})_{m/w} = 3$ , for each of the single point

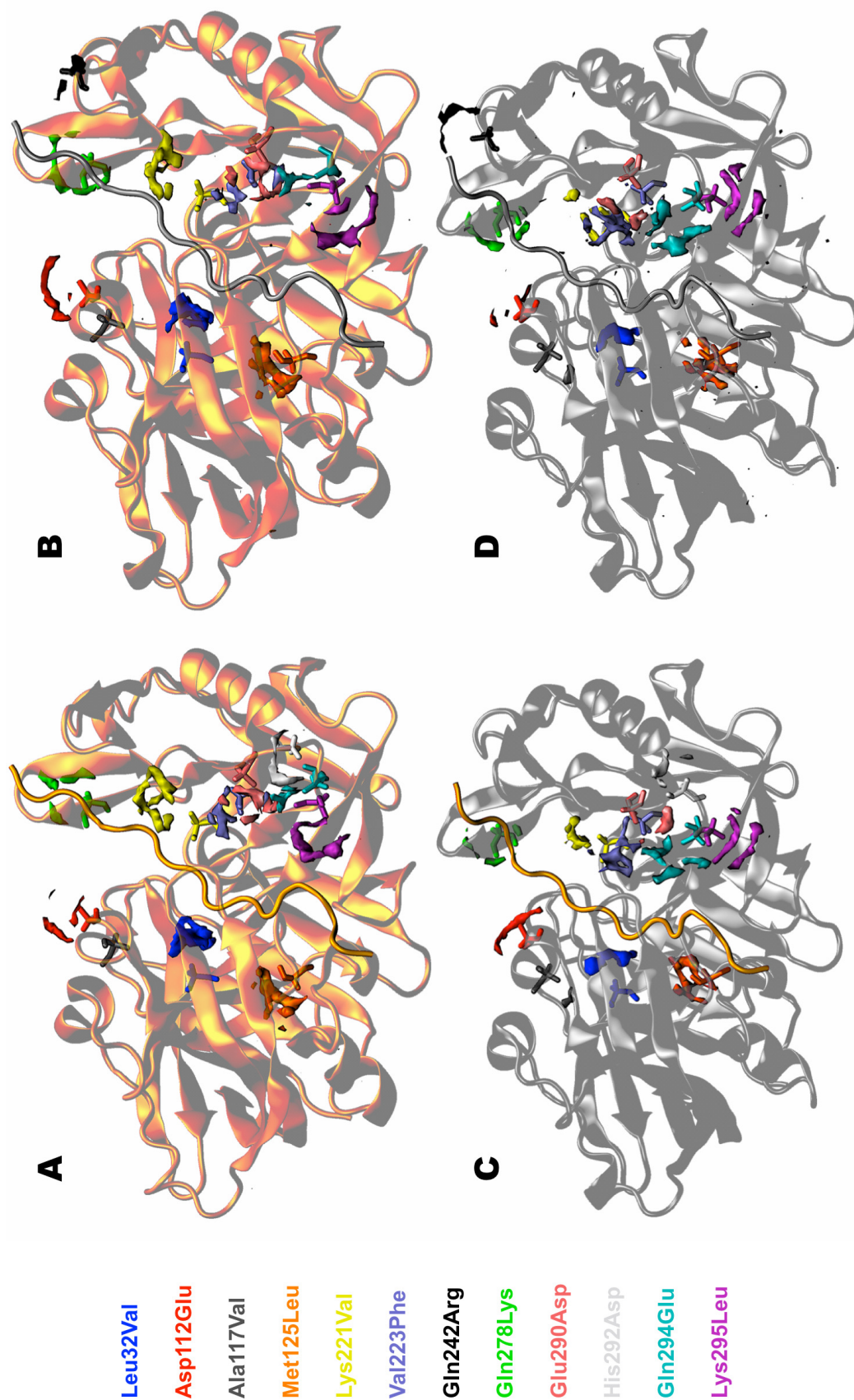


Figure 14: Illustration of the changes in local solvent density for 12 single-point mutations on the four different complexes used in this study. **A** = Bovine chymosin (gold) in complex with bovine  $\kappa$ -casein (gold). **B** = Bovine chymosin (gold) in complex with camel  $\kappa$ -casein (silver). **C** = Camel chymosin (silver) in complex with bovine  $\kappa$ -casein (gold). **D** = Camel chymosin (silver) in complex with camel  $\kappa$ -casein (silver). Each of the colored surfaces corresponds to an isosurface of a residue on which  $g(\mathbf{r})_{\text{mutant}} - g(\mathbf{r})_{\text{wildtype}} = 3$ . The colours of the isosurfaces correspond to alanine mutation of: Leu32Val (blue), Asp112Glu (red), Ala117Val (grey), Met125Leu (orange), Lys221Val (yellow), Val223Phe (lilac), Gln242Arg (black), Gln278Lys (green), Glu290Asp (pink), His292Asp (off-white), Gln294Glu (cyan) and Lys295Leu (purple).

mutations introduced in the alanine scanning experiments in the four complexes. Changes in the local solvation are observed around each single-point mutation. As would be expected, the largest changes in solvation are localised within the binding site, with the most significant changes occurring due to changes in excluded volume.

## 6 Conclusions

Using molecular dynamics simulations and free energy calculations, we have investigated binding in four different chymosin- $\kappa$ -casein complexes (Bov/Bov, Bov/Cam, Cam/Bov, Cam/Cam). By way of computational alanine scanning calculations, we have identified the influence that differences in the primary sequence of bovine and camel chymosin ("natural mutations") have on the binding thermodynamics in these complexes. Four of the natural mutations investigated here do not appear to differ in their contribution to  $\Delta\Delta G_{bind}$  as both the bovine and camel variants produce similar alanine scanning results. It is worth remembering that these residues can have specific interactions that assist in forming the optimal orientation of the complex, or facilitate the correct binding of  $\kappa$ -casein to chymosin. The alanine scanning results shows that there are eight important residues (112, 221, 242, 278, 290, 292, 294, and 295) that selectively influence binding thermodynamics. For Gln242Arg and Gln278Lys, the residue in bovine chymosin is more energetically favourable for binding. In the camel chymosin systems the alanine mutations are energetically favoured suggesting the polar positive residues in camel chymosin adversely influences the binding thermodynamics with  $\kappa$ -casein. By contrast, for residues Asp112Glu, Lys221Val and Lys295Leu, the native camel variant is most favoured. All of these residues occupy separate and predominantly non-polar pockets along the binding cleft where the natural polar positive residues in bovine chymosin adversely influences the binding thermodynamics. Analysis of the solvent density distributions obtained by 3DRISM illustrate that, as might be expected, mutation of binding site residues to alanine leads to localised changes in solvent density, with the largest contributions coming from excluded volume effects and polar functional groups.

It should be noted that there are a number of factors that are a part of the enzymatic process, and binding free energy is just one of them. Factors not considered in this study such as covalent bond breaking/forming and association/dissociation kinetics also affect the enzymatic process. Nonetheless, on the basis of the analysis carried out here, several residues have been identified for mutation with the aim of selectively modifying the bind free energy. Other aspects of chymosin catalysis, including the enzymatic reaction mechanism, are the subject of ongoing investigation in our research groups.

## Acknowledgments

We thank Maksim Mišin for implementing the PC and PC+ functional in a modified version of the MMPBSA.py program. S.M.A and D.S.P. are grateful for use of the EPSRC funded ARCHIE-WeSt High Performance Computer ([www.archie-west.ac.uk](http://www.archie-west.ac.uk), EPSRC Grant No. EP/K000586/1). D.S.P. thanks the University of Strathclyde for support through its Strategic Appointment and Investment Scheme. J.S. and B.S. are grateful for us of the Grendel hybrid CPU/GPU cluster at the Centre for Scientific Computing Aarhus ([www.cscaa.dk](http://www.cscaa.dk)). J.S. thanks the Benzon foundation for funding.

## References

- [1] Foltmann B., Pedersen V. B., Kauffman D., Wybrandt G.. The primary structure of calf chymosin. *J. Biol. Chem.*. 1979;254(17):8447-8456.
- [2] Kappeler S. R., Brink H. J., Rahbek-Nielsen H., et al. Characterization of recombinant camel chymosin reveals superior properties for the coagulation of bovine and camel milk. *Biochem. Biophys. Res. Commun.*. 2006;342(2):647-654.
- [3] Palmer D. S., Christensen A. U., Sørensen J., Celik L., Qvist K. Bruun, Schiøtt B.. Bovine Chymosin: A Computational Study of Recognition and Binding of Bovine  $\kappa$ -Casein. *Biochemistry*. 2010;49(11):2563-2573.

- [4] Pearl L., Blundell T.. The active site of aspartic proteinases. *FEBS lett.*. 1984;174(1):96-101.
- [5] Newman M., Watson F., Roychowdhury P., et al. X-ray analyses of aspartic proteinase: V structure and refinement at 2.0Å resolution of the aspartic proteinase from *Mucor pusillus*. *J. Mol. Biol.*. 1993;(230):260–283.
- [6] Cooper J. B., Khan G., Taylor G., Tickle I. J., Blundell T. L.. X-ray analyses of aspartic proteinases II: Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3Å. *J. Mol. Biol.*. 1990;3(214):199–222.
- [7] Gustchina E., Rumsh L., Ginodman L., Majer P., Andreeva N.. Post X-ray crystallographic studies of chymosin: the existence of two structural forms and the regulation of activity by the interaction with the histidine-proline cluster of kappa-casein. *FEBS Lett.*. 1996;379(1):60-62.
- [8] Ansari S.M., Coletta A., Skeby K. K., Sørensen J., Schiøtt B., Palmer D. S.. Allosteric-Activation Mechanism of Bovine Chymosin Revealed by Bias-Exchange Metadynamics and Molecular Dynamics Simulations. *J. Phys. Chem. B*. 2016;120(40):10453-10462.
- [9] Jensen J., Molgaard A., Poulsen J., et al. Structural comparison of the milk-clotting enzymes bovine and camel chymosin. *Acta Cryst. Sect. A*. 2011;67(a1):C772–C772.
- [10] Jensen J. Langholm, Mølgaard A., Poulsen J-C. Navarro, et al. Camel and bovine chymosin: the relationship between their structures and cheese-making properties. *Acta Crystallogr. Sect. D-Biol. Crystallogr.*. 2013;69(5):901–913.
- [11] Schechter I., Berger A.. On the size of the active site in proteases. I. Papain.. *Biochem. Biophys. Res. Commun.*. 1967;27:157–162.
- [12] Dalglish D. G.. Dalglish, D. G.; Casein Micelles as Colloids: Surface Structures and Stabilities.. *J. Dairy Sci.*. 1998;81:3013-3018.

- [13] Groves M. R., Dhanaraj V., Badasso M., et al. A 2.3 Å resolution structure of chymosin complexed with a reduced bond inhibitor shows that the active site beta-hairpin flap is rearranged when compared with the native crystal structure. *Protein Eng. Des. Sel.*. 1998;11(10):833-840.
- [14] Plowman J. E., Creamer L. K.. Restrained molecular dynamics study of the interaction between bovine kappa-casein peptide 98-111 and bovine chymosin and porcine pepsin. *J. Dairy Res.*. 1995;62(3):451-467.
- [15] Plowman J. E., Creamer L. K., Smith M. H., Hill J. P.. Restrained molecular dynamics investigation of the differences in association of chymosin to k-caseins A and C. *J. Dairy Res.*. 1997;64:299-304.
- [16] Plowman J. E., Smith M. H., Creamer L. K., et al. Proton assignment and structural features of a peptide from the chymosin-sensitive region of bovine k-casein determined by 2D-NMR spectroscopy. *Magn. Reson. Chem.*. 1994;32:458-464.
- [17] Bairoch A., Apweiler R., Wu C. H., et al. The Universal Protein Resource (UniProt). *Nucl. Acids Res.*. 2005;33(suppl 1):154 - 159.
- [18] Macheboef D., Coulon J. B., D'Hour P.. Effect of breed, protein genetic variants and feeding on cows' milk coagulation properties.. *J. Dairy Res.*. 1993;9:373-374.
- [19] Post A.. *Fractionation of bovine casein and enrichment of functional casein peptides*. Munchen: Dr. Hut.; 2012.
- [20] Doi H., Park B., Ibuki F., Kanamori M.. Heterogeneity and composition of K-casein from bovine colostrum. *Agr. Bio. Chem.*. 1980;44(4):813–820.
- [21] Palmer D., Sørensen J., Schiøtt B., Fedorov M.. Solvent Binding Analysis and Computational Alanine Scanning of the Bovine Chymosin-Bovine  $\kappa$ -Casein Complex Using Molecular Integral Equation Theory. *J. Chem. Theory Comput.*. 2013;9(12):5706–5717.

- [22] Jorgensen W. L.. The many roles of computation in drug discovery. *Science*. 2004;303(5665):1813–1818.
- [23] Aqvist J., Medina C., Samuelsson J. E.. A New Method For Predicting Binding-Affinity In Computer-Aided Drug Design. *Protein Eng.*. 1994;7(3):385–391.
- [24] Srinivasan J., Cheatham T. E., Cieplak P., Kollman P. A., Case D. A.. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.*. 1998;120(37):9401–9409.
- [25] Kollman P. A., Massova I., Reyes C., et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*. 2000;33(12):889–897.
- [26] Baron R., Setny P., McCammon J. A.. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.*. 2010;132(34):12091-12097.
- [27] Genheden S., Luchko T., Gusarov S., Kovalenko A., Ryde U.. An MM/3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B*. 2010;114:8505–8516.
- [28] Jensen F., Palmer D. S.. Harmonic Vibrational Analysis in Delocalized Internal Coordinates. *J. Chem. Theory Comput.*. 2011;7(1):223–230.
- [29] Palmer D. S., Sergiievskyi V. P, Jensen F., Fedorov M. V.. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. *J. Chem. Phys.*. 2010;133(4):044104.
- [30] Palmer D. S., Frolov A. I., Ratkova E. L., Fedorov M. V.. Toward a Universal Model To Calculate the Solvation Thermodynamics of Druglike Molecules: The Importance of New Experimental Databases. *Mol. Pharmaceutics*. 2011;8(4):1423–1429.
- [31] Misin M., Fedorov M. V., Palmer D. S.. Accurate hydration free energies at a wide range of temperatures from 3D-RISM. *J. Chem. Phys.*. 2015;142(9):091105.

- [32] Misin M., Fedorov M. V., Palmer D. S.. Hydration free energies of molecular ions from theory and simulation. *J. Phys. Chem. B.* 2016;120(5):975–983.
- [33] Misin M., Vainikka P. A., Fedorov M. V., Palmer D. S.. Salting-out effects by pressure-corrected 3D-RISM. *J. Chem. Phys.*. 2016;145(19):194501.
- [34] Beglov D., Roux B.. An integral equation to describe the solvation of polar molecules in liquid water. *J. Phys. Chem.*. 1997;101:7821–7826.
- [35] Du Q. H., Beglov D., Roux B.. Solvation free energy of polar and nonpolar molecules in water: An extended interaction site integral equation theory in three dimensions. *J. Phys. Chem. B.* 2000;104(4):796–805.
- [36] Hirata F., ed. *Molecular theory of solvation*. Kluwer Academic Publishers, Dordrecht, Netherlands; 2003.
- [37] Imai T., Oda K., Kovalenko A., Hirata F., Kidera A.. Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design. *J. Am. Chem. Soc.*. 2009;131(34):12430–12440.
- [38] Ratkova E. L., Palmer D. S., Fedorov M. V.. Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem. Rev.*. 2015;115(13):6312–6356.
- [39] Luchko T., Gusarov S., Roe D. R., et al. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.*. 2010;6(3):607–624.
- [40] Hansen J.-P., McDonald I. R.. *Theory of Simple Liquids, 4th ed.* Academic Press; 2000.
- [41] Duh D. M., Haymet A. D. J.. Integral-Equation Theory For Uncharged Liquids: The Lennard-Jones Fluid And The Bridge Function. *J. Chem. Phys.*. 1995;103(7):2625–2633.



- [42] Chuev G. N., Vyalov I., Georgi N.. Extraction of atom-atom bridge and direct correlation functions from molecular simulations: A test for ambient water. *Chem. Phys. Lett.*. 2013;561-562(0):175-178.
- [43] Kast S. M., Kloss T.. Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. *J. Chem. Phys.*. 2008;129(23):236101.
- [44] Kovalenko A., Hirata F.. Potential of mean force between two molecular ions in a polar molecular solvent: A study by the three-dimensional reference interaction site model. *J. Phys. Chem. B.* 1999;103:7942–7957.
- [45] Ratkova E. L., Chuev G. N., Sergiievskiy V. P., Fedorov M. V.. An Accurate Prediction of Hydration Free Energies by Combination of Molecular Integral Equations Theory with Structural Descriptors. *J. Phys. Chem. B.* 2010;114(37):12068–12079.
- [46] Ten-no S., Jung J., Chuman H., Kawashima Y.. Assessment of free energy expressions in RISM integral equation theory: theoretical prediction of partition coefficients revisited. *Mol. Phys.*. 2010;108(3-4):327–332.
- [47] Chandler D., Singh Y., Richardson D. M.. Excess Electrons In Simple Fluids .1. General Equilibrium-Theory For Classical Hard-Sphere Solvents. *J. Chem. Phys.*. 1984;81(4):1975–1982.
- [48] Sergiievskiy V. P., Jeanmairet G., Levesque M., Borgis D.. Fast computation of solvation free energies with molecular density functional theory: thermodynamic-ensemble partial molar volume corrections. *J. Phys. Chem. Lett.*. 2014;5(11):1935–1942.
- [49] Misin M., Palmer D. S., Fedorov M. V.. Predicting solvation free energies using parameter-free solvent models. *J. Phys. Chem. B.* 2016;(25):5724-5731.

- [50] Li B., Matveev A. V., Rösch N.. Three-dimensional reference interaction site model solvent combined with a quantum mechanical treatment of the solute. *Comp. Theor. Chem.*. 2015;1070:143–151.
- [51] Sørensen J., Palmer D. S., Qvist K. B., Schiøtt B.. Initial Stage of Cheese Production: A Molecular Modeling Study of Bovine and Camel Chymosin Complexed with Peptides from the Chymosin-Sensitive Region of kappa-Casein. *J. Agric. Food. Chem.*. 2011;59(10):5636–5647.
- [52] Phillips J. C., Braun R., Wang W., et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*. 2005;26(16):1781–1802.
- [53] Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., Klein M. L.. Comparison Of Simple Potential Functions For Simulating Liquid Water. *J. Chem. Phys.*. 1983;79(2):926–935.
- [54] Hoover W. G.. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A.* 1985;31(3):1695–1697.
- [55] Nose S., Klein M. L.. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*. 1983;50(5):1055-1076.
- [56] Martyna G. J., Tobias D. J., Klein M. L.. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*. 1994;101:4177-4189.
- [57] Ewald P. P.. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*. 1921;64:253-287.
- [58] Darden T., York D., Pedersen L.. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.*. 1993;98:10089–10092.
- [59] York D. M., Wlodawer A., Pedersen L. G., Darden T. A.. Atomic-level accuracy in simulations of large protein crystals. *Proc. Natl. Acad. Sci. U. S. A.*. 1994;91(18):8715-8718.

- [60] Ryckaert J-P., Ciccotti G., Berendsen H. J. C.. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comp. Phys.*. 1977;23:327–341.
- [61] Weinbach Y., Elber R.. Revisiting and parallelizing SHAKE. *J. Comput. Phys.*. 2005;209:193-206.
- [62] Blinov N., Dorosh L., Wishart D., Kovalenko A.. Association Thermodynamics and Conformational Stability of beta-Sheet Amyloid beta(17-42) Oligomers: Effects of E22Q (Dutch) Mutation and Charge Neutralization. *Biophys. J.*. 2010;98(2):282–296.
- [63] Case D. A., Berryman J. T., Betz R. M., et al. *AMBER*. 2015.
- [64] Gohlke H., Case D. A.. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.*. 2004;25(2):238–250.
- [65] Palmer D., Frolov A., Ratkova E., Fedorov M.. Towards a universal method for calculating hydration free energies: a 3D reference interaction site model with partial molar volume correction. *J. Phys.*. 2010;22(49):492101.
- [66] Kovalenko A., Ten-No S., Hirata F.. Solution of three-dimensional reference interaction site model and hypernetted chain equations for simple point charge water by modified method of direct inversion in iterative subspace. *J. Comput. Chem.*. 1999;20(9):928–936.
- [67] Massova I., Kollman P. A.. Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.*. 1999;121(36):8133-8143.
- [68] Bradshaw R. T., Patel B.H., Tate E.W., Leatherbarrow R.J., Gould I.R. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. *Protein Eng. Des. Sel.*. 2011;24:197-207.

- [69] Strop P., Sedlacek J., Stys J., et al. Engineering enzyme subsite specificity: preparation, kinetic characterization, and X-ray analysis at 2.0-Å resolution of Val111Phe site-mutated calf chymosin. *Biochemistry*. 1990;29(42):9863-9871.
- [70] Huo S., Massova I., Kollman P.A.. Computational alanine scanning of the 1 : 1 human growth hormone-receptor complex.. *J. Comput. Chem.*. 2002;23:15-27.
- [71] Moreira I.S., Fernandes P.A., Ramos M.J. Computational alanine scanning mutagenesis - An improved methodological approach. *J. Comput. Chem.*. 2007;28:644-654.
- [72] Sørensen J., Palmer D. S., Schiøtt B.. Hot-Spot Mapping of the Interactions between Chymosin and Bovine kappa-Casein. *J. Agr. Food Chem.*. 2013;61:7949–7959.
- [73] Visser S., Rooijen P. J. Van, Schattenkerk C., Kerling K. E.. Peptide substrates for chymosin (rennin). Kinetic studies with bovine kappa-casein-(103-108)-hexapeptide analogues. *Biochimica et biophysica acta*. 1977;481(1):171-176.
- [74] White A. D., Keefe A. J., Ella-Menye J.-R., et al. Free Energy of Solvated Salt Bridges: A Simulation and Experimental Study. *J. Phys. Chem. B*. 2013;117(24):7254-7259.
- [75] Bairoch A., Apweiler R., Wu C. H., et al. The Universal Protein Resource (UniProt). *Nucleic Acids Res.*. 2005;33:D154–D159.
- [76] Visser S.. Peptide substrates for chymosin (Rennin). Kinetic studies with peptides of different chain length including parts of the sequence 101-112 of bovine kappa-casein. *Biochim. et Biophys. Acta (BBA) - Enzymol.* 1976;438(1):265–272.